Faculty Working Papers

INADMISSIBILITY OF THE LIMITED INFORMATION MAXIMUM LIKELIHOOD ESTIMATOR WHEN THE DISTURBANCES ARE SMALL

Kimio Morimune

#413

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

FACULTY WORKING PAPERS

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

July 1, 1977

INADMISSIBILITY OF THE LIMITED INFORMATION MAXIMUM LIKELIHOOD ESTIMATOR WHEN THE DISTURBANCES ARE SMALL

Kimio Morimune

#413

Inadmissibility of the Limited Information Maximum Likelihood Estimator When the Disturbances are Small

Kimio Morimune*

June 30, 1977

Prepared Under

NATIONAL SCIENCE FOUNDATION GRANT SOC 76-22232

*Department of Economics, University of Illinois at Urbana-Champaign and Institute of Economic Research, Kyoto University. The author thanks Takamitsu Sawa for his valuable comments. Other helpful comments were contributed by Theodore W. Anderson, Joseph B. Kadane and Arnold Zellner. He also wishes to thank Andrew J. Buck for improving the style of the paper.

Abstract

The Stein Paradox is extended to the estimation of a structural form equation, the Limited Information Maximum Likelihood estimator is shown to be inadmissible. Instead of Stein's estimator, we propose estimators combining linearly the Limited Information Maximum Likelihood estimator and the Two Stage or the Ordinary Least Squares estimator.

1. Introduction

In a multiple regression problem in which the dependent variable and (3 or more) independent variables have a joint normal distribution, Stein [1960] and Baranchik [1973] established the inadmissibility of the maximum likelihood estimator. This result is extended to the estimation of a single structural form equation: the Limited Information Maximum Likelihood (LIML) estimator will be shown to be inadmissible. Instead of Stein's estimator, we will propose estimators which are linear combinations of the LIML estimator and the Two Stage Least Squres (TSLS) or the Ordinary Least Squares (OLS) estimator. The method of analysis is the small-σ approach first introduced by Kadane [1971].

The theoretical research into the econometrics of simultaneous equations was greatly enhanced by the discovery of the fact that OLS gives inconsistent estimates. Since then, several consistent estimators have been proposed. The LIML and the TSLS estimators are the most successful ones.

It is only in this decade that the exact distributions and moments of the above estimators have been obtained, especially for the case of two endogenous variables. Two of the most interesting findings in the small-sample analyses are: LIML does not have any exact moments, and the small-sample bias of TSLS is more serious than was expected. Although the bias of TSLS is smaller than that of OLS, this fact naturally motivated researchers to devise estimators which are less biased than TSLS. Nagar's unbiased k-class estimator [1959] was a revealed effort in this direction. Sawa [1973] considered improving TSLS by combining it with OLS.

However, it has recently been found that LIML may possibly have a smaller bias and mean square error than TSLS when the number of exogenous variables excluded from the structural equation is large (such as in large econometric models). Anderson [1974] showed this for a structural equation with two endogenous variables. Kadane [1971] showed this fact with respect to bias by using his small-σ approach (reviewed in Lemma 3.1). Further, in the numerical computations of the distributions of TSLS, Anderson and Sawa [1977] have shown that the asymptotic normality of TSLS may be more or less unrealistic in large econometric models.

The article considers some methods of eliminating the small-sample bias of the LIML estimator, and proposes improved estimators. The improved estimators are unbiased to a certain order (almost unbiased). As was done in Sawa's paper [1973], they will be called combined estimators. The combined estimators are convex linear combinations of the LIML and the k-class (fixed k) estimators. The mean square errors of the combined estimators will also be derived. Then the inadmissibility of the LIML estimator will be shown: The combined estimators dominate LIML.

In Section 2, the model and the notation will be briefly explained. Section 3 shows the derivation of "almost unbiased" combined estimators. In Section 4, the mean-square-errors (MSE) of the combined estimators will be derived by the small-σ asymptotic approach. Then a comparison of the combined estimators with LIML, TSLS, and OLS will be made. In Section 5, some Monte Carlo experiments will be reviewed from the view-point of the combined estimators. It is interesting to find that the

structural equations which have been dealt with in these Monte Carlo experiments do not really represent empirical econometric models, especially the so-called economy-wide econometric models. Finally the section will close by applying the combined estimators to Klein's Model 1 [1950].

2. Model and Notation

Let a single structural equation be

$$y_1 = Y_2\beta + Z_1\gamma + u \ , \tag{2.1}$$

where y_1 and Y_2 are T by 1 and T by G_1 matrices, respectively, of observations on the endogenous variables, Z_1 is a T by K_1 matrix of observations on the K_1 exogenous variables, β and γ are column vectors with G_1 and K_1 unknown parameters, and u is a column vector of T disturbances. The reduced form of $Y = (y\ Y_1)$ is defined as

$$Y = Z\Pi + V \tag{2.2}$$

where Z is a T by K matrix of exogenous variables (full rank), Π is a K by $(G_1 + 1)$ matrix of the reduced form coefficients, and $V = (v_1\ V_2)$ is a T by $(1 + G_1)$ matrix of disturbances. We assume the following on the reduced form disturbance term.

ASSUMPTION 1. The rows of V are independently normally distributed, each row having mean 0 and (nonsingular) covariance matrix

$$\Omega = (\omega_{ij}) = \begin{bmatrix} \omega_{11} & \omega_{12}' \\ \omega_{21} & \Omega_{22} \end{bmatrix} . \tag{2.3}$$

In order to relate (2.1) and (2.2), we partition $\underset{\sim}{\Pi}$ into K_1 and K_2 $(= K-K_1)$ rows, and into 1 and G_1 columns, respectively;

$$(2.4) \qquad \underset{\sim}{\Pi} = \begin{pmatrix} \underset{\sim}{\pi}_{11} & \underset{\sim}{\Pi}_{12} \\ \underset{\sim}{\pi}_{21} & \underset{\sim}{\Pi}_{22} \end{pmatrix} .$$

Post-multiplying $\underset{\sim}{\Pi}$ by $(1, -\underset{\sim}{\beta}')'$, we obtain (2.1). We note that

$$\underset{\sim}{u} = \underset{\sim}{v}_1 - \underset{\sim}{V}_2\underset{\sim}{\beta} ,$$

and

$$\underset{\sim}{\gamma} = \underset{\sim}{\pi}_{11} - \underset{\sim}{\Pi}_{12}\underset{\sim}{\beta} .$$

In order that (2.1) be properly written with $\underset{\sim}{Z}_2$ omitted

$$(2.5) \qquad \underset{\sim}{\pi}_{21} = \underset{\sim}{\Pi}_{22}\underset{\sim}{\beta} .$$

If the reduced form implies a unique set of structural equations, then (2.1) is identified. This is the case if (2.5) has a unique solution for $\underset{\sim}{\beta}$.

ASSUMPTION 2. The matrix $(\underset{\sim}{\pi}_{21}\ \underset{\sim}{\Pi}_{22})$ is of rank G_1 and $\underset{\sim}{\Pi}_{22}$ is also of rank G_1.

The components of $\underset{\sim}{u}$ are independently normally distributed with mean 0 and variance

$$\sigma^2 = \omega_{11} - 2\,\underset{\sim}{\omega}_{12}\underset{\sim}{\beta} + \underset{\sim}{\beta}'\underset{\sim}{\Omega}_{22}\underset{\sim}{\beta} .$$

The existence of (2.2) implies that there are at least (G_1+1) structural equations in the system. If there are more than (G_1+1) equations, there must be more endogenous variables. However, the only aspect of the entire system that concerns us is that all of the predetermined variables $\underset{\sim}{Z}$ are exogenous. Assumption 2 ensures that there are at least G_1 such variables which occur in the system and do not appear in the structural equation of interest.

For convenience, we use the notation

$$L = K_2 - G_1 \tag{2.6}$$

which is no more than the degree of over-identification of the equation (2.1).

Finally, we present some notation which is used throughout the paper.

$$\underset{\sim}{\theta} = \begin{pmatrix} \underset{\sim}{\beta} \\ \underset{\sim}{\gamma} \end{pmatrix} ,$$

$\hat{\underset{\sim}{\theta}}_{\lambda}$: The LIML estimator of $\underset{\sim}{\theta}$,

$\hat{\underset{\sim}{\theta}}_{k}$: The k-class estimator of $\underset{\sim}{\theta}$ (k any fixed constant),

$$\mathrm{MSE}(\hat{\underset{\sim}{\theta}}) = E(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta})(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta})' ,$$

$$\underset{\sim}{Q} = \begin{pmatrix} \underset{\sim}{\Pi}_2' \underset{\sim}{Z}' \underset{\sim}{Z} \underset{\sim}{\Pi}_2 & \underset{\sim}{\Pi}_2' \underset{\sim}{Z}' \underset{\sim}{Z}_1 \\ \underset{\sim}{Z}_1' \underset{\sim}{Z} \underset{\sim}{\Pi}_2 & \underset{\sim}{Z}_1' \underset{\sim}{Z}_1 \end{pmatrix}^{-1} ,$$

$$\underset{\sim}{q} = (\underset{\sim}{\omega}_{12}' - \underset{\sim}{\beta}' \underset{\sim}{\Omega}_{22}, \underset{\sim}{0}') ,$$

$$\underset{\sim}{C}_1 = \underset{\sim}{q}\, \underset{\sim}{q}' ,$$

$$\underset{\sim}{C}_2 = \begin{pmatrix} \underset{\sim}{\Omega}_{22} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{pmatrix} - \underset{\sim}{C}_1 ,$$

$$\underset{\sim}{A} = \underset{\sim}{Q}\, \underset{\sim}{C}_1\, \underset{\sim}{Q} ,$$

$$\underset{\sim}{B} = \underset{\sim}{Q}\, \underset{\sim}{C}_2\, \underset{\sim}{Q} .$$

3. Combined Estimators

Certain approximate moments of the estimators have been given by Kadane [1973]. He expanded the k-class estimator as well as the LIML estimator in a power series in terms of the standard deviation (σ) of

the disturbance in the relevant structural equation; thereby he calculated the approximate bias and mean square error matrices to a suitable order of σ. This method of expansion is called "small-σ asymptotic expansion".

Anderson [1977] makes it clear that the first two moments obtained by Kadane are those of a kind of asymptotic expansion of the distributions of the estimators.

We also note that the LIML estimator does not have any exact moments.

We first introduce Kadane's Theorem 1 [1973] as Lemma 3.1.

LEMMA 3.1. *The small-σ asymptotic biases of the LIML and the k-class estimators (k-fixed) are given by*

$$E\,[\hat{\underset{\sim}{\theta}}_{\lambda} - \underset{\sim}{\theta}] = -\,\sigma^2\, \underset{\sim}{Q}\underset{\sim}{q} + O(\sigma^4)\ ,$$

and

$$E\,[\hat{\underset{\sim}{\theta}}_{k} - \underset{\sim}{\theta}] = \sigma^2\,[(1-k)(T-K) + L-1]\, \underset{\sim}{Q}\underset{\sim}{q} + O(\sigma^4)$$

respectively.

The next theorem provides the combined estimator between the LIML and the k-class estimators which is small-σ asymptotically unbiased (k-fixed).

Theorem 3.1. The estimator

$$\hat{\underset{\sim}{\theta}}_{k\lambda} = \frac{(1-k)(T-K) + L-1}{(1-k)(T-K) + L}\,\hat{\underset{\sim}{\theta}}_{\lambda} + \frac{1}{(1-k)(T-K) + L}\,\hat{\underset{\sim}{\theta}}_{k}$$

is a small-σ asymptotically unbiased estimator of θ (k is fixed).

The proof of this theorem is direct from Lemma 3.1. The combined LIML-TSLS ($\hat{\underset{\sim}{\theta}}_{1\lambda}$) and LIML-OLS ($\hat{\underset{\sim}{\theta}}_{0\lambda}$) estimators are directly obtainable from Theorem 3.1. For convenience, we show them below explicitly:

$$\hat{\underset{\sim}{\theta}}_{1\lambda} = \frac{L-1}{L}\hat{\underset{\sim}{\theta}}_{\lambda} + \frac{1}{L}\hat{\underset{\sim}{\theta}}_{1}, \tag{3.1}$$

and

$$\hat{\underset{\sim}{\theta}}_{0\lambda} = \frac{T-K+L-1}{T-K+L}\hat{\underset{\sim}{\theta}}_{\lambda} + \frac{1}{T-K+L}\hat{\underset{\sim}{\theta}}_{0}.$$

4. Mean Square Error

The decrease in bias is usually obtained only at the cost of the increase in variance. It was the case for Sawa's estimator; the mean square error (MSE) of the combined estimator can be larger than that of TSLS. However, the MSE of our estimators ($\hat{\underset{\sim}{\theta}}_{1\lambda}$, $\hat{\underset{\sim}{\theta}}_{0\lambda}$ defined in 3.1) are always smaller than that of LIML.

Since MSE is one of the most often employed criterion for goodness of the estimators, we derive the MSE of the estimators ($\hat{\underset{\sim}{\theta}}_{1\lambda}$, $\hat{\underset{\sim}{\theta}}_{0\lambda}$) and compare them with the MSE of LIML, TSLS, OLS. Throughout this section, the method for the derivation of the MSE is the small-σ approximation.

<u>LEMMA 4.1. The MSE of $\hat{\underset{\sim}{\theta}}_{1\lambda}$ is small-σ asymptotically expanded as follows: Provided $T-K \geq 3$</u>

$$MSE(\hat{\theta}_{1\lambda}) = \sigma^2 \underset{\sim}{Q} + \sigma^4 \{tr(\underset{\sim}{C}_1\underset{\sim}{Q})\underset{\sim}{Q} + tr(\underset{\sim}{C}_2\underset{\sim}{Q})\underset{\sim}{Q} + \frac{L+2}{L}\underset{\sim}{A} + [L + \frac{(L-1)^2(L+2)}{L(T-K-2)}]\underset{\sim}{B}\}.$$

The proof will be given in the appendix.

<u>LEMMA 4.2. The MSE of $\hat{\underset{\sim}{\theta}}_{0\lambda}$ is small-σ asymptotically expanded as follows: Provided $T-K \geq 3$</u>

$$MSE(\hat{\underset{\sim}{\theta}}_{0\lambda}) = \sigma^2 \underset{\sim}{Q} + \sigma^4 \{tr(\underset{\sim}{C}_1\underset{\sim}{Q})\underset{\sim}{Q} + tr(\underset{\sim}{C}_2\underset{\sim}{Q})\underset{\sim}{Q} + \frac{T-K+L+2}{T-K+L}\underset{\sim}{A} + \frac{T-K+L-2}{T-K-2}(L + \frac{1}{T-K+L})\underset{\sim}{B}\}.$$

The proof will also be given in the appendix. Observing Lemma 4.1 and Lemma 4.2, we see that $\sigma^2\underset{\sim}{Q}$ and σ^4 $[tr(\underset{\sim}{C}_1\underset{\sim}{Q}) + tr(\underset{\sim}{C}_2\underset{\sim}{Q})]\underset{\sim}{Q}$ are common for

both the MSEs. By definition of $\underset{\sim}{C}_2$ matrix, it is easy to see;

$$\mathrm{tr}(C_1Q) + \mathrm{tr}(C_2Q) = \mathrm{tr}\,\{\underset{\sim}{\Omega}_{22}\,[\underset{\sim}{\Pi}_{22}'\,\underset{\sim}{Z}_2'\,\overline{\underset{\sim}{P}}_{Z_1}\,\underset{\sim}{Z}_2\,\underset{\sim}{\Pi}_{22}]^{-1}\}\ .$$

<u>LEMMA 4.3. If one of LIML and TSLS dominates the other, then $\hat{\underset{\sim}{\theta}}_{1\lambda}$ also dominates it. If one of LIML and OLS dominates the other, then $\hat{\underset{\sim}{\theta}}_{0\lambda}$ also dominates it.</u>

(Proof) We show this theorem for the estimator combining LIML with the k-class estimator (fixed k). For convenience we define

$$r_k = (1-k)(T-K) + L-1$$

and the deviation of an estimator from its mean is defined as

$$\hat{\underset{\sim}{e}}_\ell = \hat{\underset{\sim}{\theta}}_\ell - \underset{\sim}{\theta}\ . \tag{4.1}$$

From Theorem 3.1, we have

$$\hat{\underset{\sim}{e}}_{k\lambda} = \frac{r_k}{r_k+1}\,\hat{\underset{\sim}{e}}_\lambda + \frac{1}{r_k+1}\,\hat{\underset{\sim}{e}}_k$$

Working out some computation, we can show

$$\begin{aligned}\mathrm{MSE}(\hat{\underset{\sim}{\theta}}_{k\lambda}) &= \frac{r_k}{r_k+1}\,\mathrm{MSE}(\hat{\underset{\sim}{\theta}}_\lambda) + \frac{1}{r_k+1}\,\mathrm{MSE}(\hat{\underset{\sim}{\theta}}_k)\\ &\quad + \frac{r_k}{(r_k+1)^2}\,[2\,E(\hat{\underset{\sim}{e}}_\lambda\hat{\underset{\sim}{e}}_k) - \mathrm{MSE}(\hat{\underset{\sim}{\theta}}_\lambda) - \mathrm{MSE}(\hat{\underset{\sim}{\theta}}_k)]\ .\end{aligned}$$

However, the last term is

$$-\frac{r_k}{(r_k+1)^2}\,E\,(\underset{\sim}{e}_\lambda - \underset{\sim}{e}_k)\,(\underset{\sim}{e}_\lambda - \underset{\sim}{e}_k)'$$

which is always negative. Then, for example, if $\hat{\underset{\sim}{\theta}}_\lambda$ is dominated by $\hat{\underset{\sim}{\theta}}_k$, we rearrange the equation so as to have

$$\begin{aligned}\mathrm{MSE}(\hat{\underset{\sim}{\theta}}_{k\lambda}) &= \mathrm{MSE}(\hat{\underset{\sim}{\theta}}_\lambda) - \frac{1}{r_k+1}\,[\mathrm{MSE}(\hat{\underset{\sim}{\theta}}_\lambda) - \mathrm{MSE}(\hat{\underset{\sim}{\theta}}_k)]\\ &\quad - \frac{r_k}{(r_k+1)^2}\,E\,[(\underset{\sim}{e}_\lambda - \underset{\sim}{e}_k)\,(\underset{\sim}{e}_\lambda - \underset{\sim}{e}_k)']\ .\end{aligned} \tag{4.2}$$

where the second term is also negative semi-definite. A similar discussion holds for the case where $\underset{\sim}{\theta}_k$ is dominated by $\hat{\underset{\sim}{\theta}}_\lambda$. (QED)

It should be pointed out that this lemma is the consequence of the convex combinations used in the construction of $\hat{\underset{\sim}{\theta}}_{k\lambda}$. However, Sawa's combined estimator $\hat{\underset{\sim}{\theta}}_{10}$ does not have this property. Further, as we mentioned before, MSE is equal to variance-covariance matrix up to $O(\sigma^4)$. Now we show the inadmissibility of the LIML estimator.

<u>Theorem 4.1. $\hat{\underset{\sim}{\theta}}_{1\lambda}$ and $\hat{\underset{\sim}{\theta}}_{0\lambda}$ dominate LIML</u>

(Proof) From Lemma 4.1 and (A.1),

$$\frac{1}{\sigma^4}\{\mathrm{MSE}(\hat{\underset{\sim}{\theta}}_\lambda) - \mathrm{MSE}(\hat{\underset{\sim}{\theta}}_{1\lambda})\} = 2\,\mathrm{tr}(\underset{\sim}{C}_1\underset{\sim}{Q})\,\underset{\sim}{Q}$$
$$+ (5 - \frac{2}{L})\,\underset{\sim}{A} + [2 + \frac{(L+2)(2L-1)}{L\,(T-K-2)}]\,\underset{\sim}{B} \geq 0\ .$$

From Lemma 4.2 and (A.1),

$$\frac{1}{\sigma^4}\{\mathrm{MSE}(\hat{\underset{\sim}{\theta}}_\lambda) - \mathrm{MSE}(\hat{\underset{\sim}{\theta}}_{0\lambda})\} = 2\,\mathrm{tr}(\underset{\sim}{C}_1\underset{\sim}{Q})\,\underset{\sim}{Q}$$
$$+ (5 - \frac{2}{T-K+L})\,\underset{\sim}{A} + (1 + \frac{L}{T-K-2})\,(2 - \frac{1}{T-K+L})\,\underset{\sim}{B} \geq 0 \quad \text{(QED)}\ .$$

As in Lemmas 4.1 and 4.2, T-K is assumed to be greater than two. This condition is necessary for the existence of moments. Contrary to Stein's result, we neither have to assume that $\underset{\sim}{\theta}$ is known nor that there are three or more predictors. Further, even though the maximum likelihood estimator in the context of a linear regression is an unbiased estimator, Stein's estimator is a biased estimator. However, in the structural form estimation, the LIML, TSLS, and OLS estimators are biased and the combined estimators are unbiased estimators (small-σ).

We show three more theorems relating to Kadane's result, which are of interest.

<u>Theorem 4.2. TSLS dominates $\hat{\underset{\sim}{\theta}}_{1\lambda}$ if $L \leq 4$</u>

(Proof) From Lemma 4.1 and (A.2),

$$\frac{1}{\sigma^4}\{\mathrm{MSE}(\hat{\underset{\sim}{\theta}}_{1\lambda}) - \mathrm{MSE}(\hat{\underset{\sim}{\theta}}_{1})\} = 2\ (L-1)\ \mathrm{tr}(C_1Q)\ Q$$
$$- \frac{(L-1)^2(L-2)}{L}\ \underset{\sim}{A} + \frac{L-1}{L}\ [2\ L + \frac{(L-1)(L+2)}{T-K-2}]\ \underset{\sim}{B}\ .$$

Since the third term is positive and $\mathrm{tr}(\underset{\sim}{C}_1\underset{\sim}{Q})\ \underset{\sim}{Q} \geq \underset{\sim}{A}$ (Kadane's Lemma [1971]), the above equation is

$$\geq - \frac{L-1}{L}\ (L^2 - 5L + 2)\ .$$

This last quantity is nonnegative for $1 \leq L \leq 4$. (QED)

This corollary is interesting since Kadane [1971] showed TSLS dominates LIML if $L \leq 6$. On the other hand, we showed that $\hat{\underset{\sim}{\theta}}_{1\lambda}$ dominates LIML always. Then we reduced the upper limit of six to four by combining LIML with TSLS.

<u>Theorem 4.3. OLS dominates $\hat{\underset{\sim}{\theta}}_{0\lambda}$ if $T-K+L \leq 4$.</u>

(Proof) Define $s = T-K+L$. From Lemma 4.2 and (A.3),

$$\frac{1}{\sigma^4}\{\mathrm{MSE}(\hat{\underset{\sim}{\theta}}_{0\lambda}) - \mathrm{MSE}(\hat{\underset{\sim}{\theta}}_{0})\} = 2\ (s-1)\ \mathrm{tr}(C_1Q)\ Q$$
$$- \frac{s-2}{s}\ (s-1)^2\ \underset{\sim}{A} + \frac{s-2}{T-K-2}\ (L + \frac{1}{s})\ \underset{\sim}{B}\ .$$

The last term is positive for positive integer of s ($T-K \geq 3$). Using Kadane's lemma again, we have

$$\geq - \frac{s-1}{s}\ (s^2 - 5s + 2)\ .$$

It should also be pointed out that $\hat{\underset{\sim}{\theta}}_{0\lambda}$ is identically equal to OLS if (T-K+L) is unity. (QED)

We have that $\hat{\underset{\sim}{\theta}}_{0\lambda}$ always dominates LIML from Theorem 4.1. Then it is natural to think that OLS dominates LIML for $T-K+L \leq 4$. This is true and we provide the next corollary which is not shown in Kadane [1971].

<u>Theorem 4.4. OLS dominates LIML if $T-K+L \leq 6$.</u>

(Proof) Define $s = T-K+L$. By (A.1) and (A.3),

$$\frac{1}{\sigma^4}\{MSE(\hat{\underset{\sim}{\theta}}_{\lambda}) - MSE(\hat{\underset{\sim}{\theta}}_{0})\} = 2\ s\ tr(C_1Q)\ Q$$

$$- (s^2 - 4s)\ \underset{\sim}{A} + (s-2)\ (1 + \frac{L+2}{T-K-2})\ \underset{\sim}{B}\ .$$

The last term is nonnegative for nonnegative s (s is nonnegative since $T \geqq K$). Then, using Kadane's lemma again, we have

$$\geqq - (s^2 - 6s)\ \underset{\sim}{A}\ . \qquad \text{(QED)}$$

Kadane showed that TSLS dominates LIML if $L \leq 6$. We showed that OLS dominates LIML if $(T-K+L) \leq 6$. Further we showed that TSLS dominates $\hat{\underset{\sim}{\theta}}_{1\lambda}$ if $L \leq 4$, and also OLS dominates $\hat{\underset{\sim}{\theta}}_{0\lambda}$ if $T-K+L \leq 4$. Even though Theorems 4.3 and 4.4 are not practically meaningful, it is interesting to find these two numbers: four and six, in all these inequalities.

5. Discussion and an Example

From Lemma 3.1, it is easy to see that the bias of LIML is independent of L, contrary to the fact that the biases of TSLS and OLS increase with L. Since the degree of over-identification is quite large in large econometric models (so-called economy wide econometric

models), TSLS and OLS may give worse estimates than LIML with respect to bias (Since T-K is also small in large models, the large sample theory may not be applicable; the consistency of LIML and TSLS is not a good base for the selection of an estimator). Reflecting this phenomena, our combined estimators are very close to LIML when L is large. On the other hand, for a small econometric model such as Klein's Model-1, (Klein [1950]), we expect some improvement in the reduction of bias (L = 4 for all three behavioral equations in Klein's Model-1).

It should also be pointed out that, in Summer's Monte Carlo experiments [1965], L is unity for the two structural equations. In this case, $\hat{\theta}_{1\lambda}$ is TSLS or the small-σ bias of TSLS is zero. In Basmann's Monte Carlo experiments [1958] (explained in Goldberger [1964], p. 361), L is also unity and the small-σ bias of TSLS is zero. Therefore it was not unusual to have found the small bias of TSLS compared with LIML and OLS in these experiments. Since Basmann's result is of special interest, we reproduce a table for convenience (from page 362 of Goldberger).

Basmann's Monte Carlo Experiment
(Reprinted from Goldberger [1964])

Bias	$y_1 =$	$-2.00\ y_2$	$+ 1.50\ y_3$	$+ 3.00\ x_1$	$- 0.60\ x_5$	$+ 10$
OLS		1.11	-0.37	-0.67	0.01	-20.69
TSLS		0.06	-0.03	-0.05	-0.05	5.28
LIML		-0.69	0.16	0.53	-0.03	-14.00

Note: T = 16, K = 6. The entries in the table are mean-biases of estimated coefficients for 200 experiments.

In this table, the mean-biases of TSLS are conspicuously smaller than those of LIML and, then OLS. This result supports our theoretical conjecture.

In Cragg's Monte Carlo experiment [1967], L is two for all of the three structural equations considered. When L is two, the bias of TSLS is the same as LIML with the opposite sign. Then we expect that TSLS and LIML will not be distinguishable from the viewpoint of bias. Actually Cragg states:

"TSLS....LIML....had rank-totals close to each other" (page 95). The mean of TSLS and LIML would have done better than them.

On the whole, the models which have been used in Monte Carlo experiments do not really represent models which are most frequently being used in empirical world: Large econometric models.

To close this section, we apply our combined estimators (3.1) and Sawa's combined estimator to Klein's Model-1. The following is Sawa's estimator combining OLS and TSLS:

$$(5.1) \qquad \hat{\underset{\sim}{\theta}}_{10} = \frac{T-K+L-1}{T-K}\,\hat{\underset{\sim}{\theta}}(1) - \frac{L-1}{T-K}\,\hat{\underset{\sim}{\theta}}(0) \ .$$

This estimator is also identical to TSLS if L = 1. The following is the result of estimation. In each vector, the order of estimates is OLS, TSLS, LIML, $\hat{\underset{\sim}{\theta}}_{10}$, $\hat{\underset{\sim}{\theta}}_{1\lambda}$, and $\hat{\underset{\sim}{\theta}}_{0\lambda}$. $\hat{\underset{\sim}{\theta}}_{10}$ is calculated by Sawa in his book [1973].

<u>Consumption Function</u>

$$C = \begin{pmatrix} 16.2366 \\ 16.5548 \\ 17.1477 \\ 16.6282 \\ 16.9995 \\ 17.0941 \end{pmatrix} + \begin{pmatrix} 0.1929 \\ 0.0173 \\ -0.2225 \\ -0.0232 \\ -0.1626 \\ -0.1981 \end{pmatrix} P + \begin{pmatrix} 0.0899 \\ 0.2162 \\ 0.3960 \\ 0.2453 \\ 0.3511 \\ 0.3780 \end{pmatrix} P_{-1} + \begin{pmatrix} 0.7962 \\ 0.8102 \\ 0.8226 \\ 0.8134 \\ 0.8195 \\ 0.8210 \end{pmatrix} (W + W')$$

Investment Function

$$I = \begin{pmatrix} 10.1258 \\ 20.2782 \\ 22.5908 \\ 22.6211 \\ 22.0127 \\ 21.8576 \end{pmatrix} + \begin{pmatrix} 0.4796 \\ 0.1502 \\ 0.0752 \\ 0.0742 \\ 0.0940 \\ 0.0990 \end{pmatrix} P + \begin{pmatrix} 0.3330 \\ 0.6159 \\ 0.6804 \\ 0.6812 \\ 0.6643 \\ 0.6600 \end{pmatrix} P_{-1} + \begin{pmatrix} 0.1118 \\ -0.1578 \\ -0.1683 \\ -0.1684 \\ -0.1657 \\ -0.1518 \end{pmatrix} K_{-1}$$

Wage Function

$$W = \begin{pmatrix} 1.4970 \\ 1.5003 \\ 1.5262 \\ 1.5011 \\ 1.5197 \\ 1.5245 \end{pmatrix} + \begin{pmatrix} 0.4395 \\ 0.4389 \\ 0.4339 \\ 0.4388 \\ 0.4352 \\ 0.4342 \end{pmatrix} E + \begin{pmatrix} 0.1461 \\ 0.1467 \\ 0.1513 \\ 0.1468 \\ 0.1502 \\ 0.1510 \end{pmatrix} E_{-1} + \begin{pmatrix} 0.1303 \\ 0.1304 \\ 0.1316 \\ 0.1304 \\ 0.1313 \\ 0.1315 \end{pmatrix} A$$

The variables are:

C: Consumption

P: Profits

W: Wage bill paid by private industry

W': Government wage bill

I: Investment

K: Capital stock

E: Total production of private industry

A: Time trend

Here $T = 21$, $K = 8$, and $L = 4$ for all the equations. All the combined estimators give rather close values to LIML. This phenomenon will occur more often in economy-wide econometric models where $(T-K)$ is very small and L is large. Since $|OLS| \leq |TSLS| \leq |LIML|$ for each coefficient, our estimators take values between OLS and LIML. However, Sawa's estimator sometimes "overshoots" LIML (bigger than $|LIML|$), especially in the Investment Function. This will also occur more frequently in large econometric models.

Appendix

In this appendix we first show Kadane's results [1971] which were used in the proofs of some corollaries. Provided $T-K \geq 3$,

(A.1) $$MSE(\hat{\theta}(\lambda)) = \sigma^2 Q + \sigma^4 [3 \, tr(C_1 Q) Q + tr(C_2 Q) Q + 6 A + \frac{(L+2)(T-K+L-2)}{T-K-2} B] + 0(\sigma^5) ,$$

(A.2) $$MSE(\hat{\theta}(1)) = \sigma^2 Q + \sigma^4 [(3-2L) \, tr(C_1 Q) Q + tr(C_2 Q) Q + ((L-2)^2 + 2) A + (2-L) B] + 0(\sigma^6) ,$$

(A.3) $$MSE(\hat{\theta}(0)) = \sigma^2 Q + \sigma^4 [(3-2(T-K+L)) \, tr(C_1 Q) Q + tr(C_2 Q) Q \\ ((T-K+L-2)^2 + 2) A + (2 - T+K-L) B] + 0(\sigma^6) .$$

Now we give the proofs of Lemmas 4.1 and 4.2. For convenience of expansions, we rewrite (2.1), (2.2) as follows:

$$y_1 = Y_1 \beta + Z_1 \gamma + \sigma u ,$$

$$(y_1 Y_2) = Z (\pi_1 \; \Pi_2) + \sigma (v_1 \; V_2) ,$$

where the disturbance term is normalized so as components of u to have unit variances.

From (4.1) and (4.2), we need to calculate

(A.4) $$E [(e_\lambda - e_k)(e_\lambda - e_k)'] .$$

However, by Kadane's Lemma A.1, we have

(A.5) $$e_\lambda - e_k = \sigma^2 Q (\dot{V} N_\lambda u - \dot{V} N_k u) + 0(\sigma^3)$$

where

$$(A.6) \qquad \underset{\sim}{V} = (\underset{\sim}{V}_2 \ \underset{\sim}{0}) \ ,$$

$$\underset{\sim}{0}: \ T \times K_1 \ ,$$

$$\underset{\sim}{N}_\lambda = \underset{\sim}{I} - \lambda \underset{\sim}{\overline{P}}_Z \ ,$$

$$\underset{\sim}{N}_k = \underset{\sim}{I} - k \underset{\sim}{\overline{P}}_Z \ ,$$

$$\underset{\sim}{\overline{P}}_Z = I - \underset{\sim}{P}_Z \ ,$$

$$\underset{\sim}{P}_Z = \underset{\sim}{Z} \ (\underset{\sim}{Z} \ \underset{\sim}{Z})^{-1} \underset{\sim}{Z}' \ .$$

Substituting (A.5) into (A.4), we have

$$(A.7) \qquad E \ [(\underset{\sim}{e}_\lambda - \underset{\sim}{e}_k)(\underset{\sim}{e}_\lambda - \underset{\sim}{e}_k)'] = \sigma^4 \ \{\underset{\sim}{Q} \ E \ [\dot{\underset{\sim}{V}}\underset{\sim}{N}_\lambda \underset{\sim}{u}\underset{\sim}{u}'\underset{\sim}{N}_\lambda \underset{\sim}{V} \ ] \ \underset{\sim}{Q}$$

$$+ \ \underset{\sim}{Q} \ E \ [\dot{\underset{\sim}{V}}\underset{\sim}{N}_k \underset{\sim}{u}\underset{\sim}{u}'\underset{\sim}{N}_k \underset{\sim}{V}] \ \underset{\sim}{Q} - 2 \ \underset{\sim}{Q} \ E \ [\dot{\underset{\sim}{V}}\underset{\sim}{N}_\lambda \underset{\sim}{u}\underset{\sim}{u}'\underset{\sim}{N}_k \underset{\sim}{V}] \ \underset{\sim}{Q}\} + O(\sigma^5) \ .$$

From Kadane's (A15), we have

$$(A.8) \qquad E \begin{Bmatrix} \dot{\underset{\sim}{V}}\underset{\sim}{N}_k \underset{\sim}{u}\underset{\sim}{u}'\underset{\sim}{N}_k \underset{\sim}{V} \\ \\ \dot{\underset{\sim}{V}}\underset{\sim}{N}_\lambda \underset{\sim}{u}\underset{\sim}{u}'\underset{\sim}{N}_\lambda \underset{\sim}{V} \end{Bmatrix} = \begin{Bmatrix} K + (k-1)^2 \ P \\ \\ K + \dfrac{L(L+2)}{P-2} \end{Bmatrix} \ \underset{\sim}{C}_2$$

$$+ \begin{Bmatrix} K \ (K+2) + 2 \ (1-k) \ K \ P + (1-k)^2 \ P \ (P+2) \\ \\ (K-L) \ (K-L+2) \end{Bmatrix} \ \underset{\sim}{C}_1$$

where $p = T-K$. Then we need to calculate the expectation of the cross-product in (A.7). For this purpose, we decompose $\underset{\sim}{V}$ into two components.

$$\underset{\sim}{V} = \underset{\sim}{W} + \underset{\sim}{u}\underset{\sim}{q}'$$

in such a way that $\underset{\sim}{W}$ is independent of $\underset{\sim}{u}$. Now

$$E \ [\dot{\underset{\sim}{V}}\underset{\sim}{N}_k \underset{\sim}{u}\underset{\sim}{u}'\underset{\sim}{N}_\lambda \underset{\sim}{V}] = E \ [\underset{\sim}{W}'\underset{\sim}{N}_k \underset{\sim}{u}\underset{\sim}{u}'\underset{\sim}{N}_\lambda \underset{\sim}{W}] + E \ [\underset{\sim}{q}\underset{\sim}{u}'\underset{\sim}{N}_k \underset{\sim}{u}\underset{\sim}{u}'\underset{\sim}{N}_\lambda \underset{\sim}{u}\underset{\sim}{q}']$$

$$= E \ [\underset{\sim}{u}'\underset{\sim}{N}_\lambda \underset{\sim}{N}_k \underset{\sim}{u}] \ \underset{\sim}{C}_2 + E \ (\underset{\sim}{u}'\underset{\sim}{N}_k \underset{\sim}{u}\underset{\sim}{u}'\underset{\sim}{N}_\lambda \underset{\sim}{u}) \ \underset{\sim}{C}_1 \ .$$

In order to evaluate the expectation, we further introduce Kadane's result (Lemma A.3):

$$\lambda = \frac{u'\bar{P}_x u}{u'\bar{P}_Z u} + 0_p(\sigma)$$

where $\bar{P}_x = I - P_x$, $P_x = X(X'X)^{-1}X'$, and $X = (Z\Pi_2\ Z_1)$.

Then

$$E\,(u'N_\lambda N_k u) = E\,\{u'u - k\,u'\,\bar{P}_Z u - \lambda\,u'\bar{P}_Z u + k\lambda u'\bar{P}_Z u\}$$

$$= E\,\{u'u - k\,u'\,\bar{P}_Z u - u'\,\bar{P}_x\,u + k\,u'\,\bar{P}_x\,u\}$$

(A.9) $$= k\,K + (1-k)\,T + (k-1)\,(T-K+L)$$

$$= K - (1-k)\,L\ .$$

Also $$E\,(u'N_k u\ u'N_\lambda u)$$

$$= E\,(u'uu'P_x u) - k\,E\,(u'\bar{P}_Z uu'P_x u)$$

$$= (T+2)\,(K-L) - k\,(T-K)\,(K-L)$$

(A.10) $$= (K-L)\,[(1-k)\,T+2+kK]\ .$$

Using (A.9) and (A.10), we have

(A.11) $$E\,[VN_\lambda uu'N_k V] = \{K - (1-k)\,L\}\,C_2 + (K-L)\,\{(1-k)\,T+2+kK\}\,C_1$$

Since the calculation of (A.7) for the combined estimator for an arbitrary k ($\hat{\theta}_{k\lambda}$) is too complicated, we derive it only for k zero and unity. For this purpose, we need to substitute k for unity or zero. Substituting unity for k in (A.8) and (A.11), we have,

(A.12) $$E\,[(e_\lambda - e_1)\,(e_\lambda - e_1)'] = L(L+2)\,A + \frac{L(L+2)}{T-K-2}\,B\ .$$

Also, for k zero, we have

$$\text{(A.13)} \qquad E\,[(\underset{\sim}{e}_\lambda - \underset{\sim}{e}_0)(\underset{\sim}{e}_\lambda - \underset{\sim}{e}_0)'] = (T-K+L)\,(T-K+L+2)\,\underset{\sim}{A}$$

$$+ (T-K+L)\,\frac{T-K+L-2}{T-K-2}\,\underset{\sim}{B} \quad .$$

Therefore, using (A.1), (A.2), (A.3), (A.12), (A.13) and (4.2), we have Lemmas 4.1 and 4.2.

References

Anderson, T. W. (1977) "Asymptotic Expansions of the Distributions of Estimates in Simultaneous Equations for Alternative Sequences", Econometrica, May.

Anderson, T. W. (1974) "An Asymptotic Expansion of the Distribution of the LIML Estimate of a Coefficient in a Simultaneous Equation System", J.A.S.A., 69, 565-572.

Anderson, T. W. and T. Sawa (1977) "Numerical Computations of the Exact and Approximate Distribution of Two-Stage Least Squares Estimate", incompleted mimeo.

Baranchik, A. J. (1973) "Inadmissibility of Maximum Likelihood Estimators in Some Multiple Regression Problems with Three or More Independent Variables", Annals of Statistics, Vol. 7, No. 2, 312-321.

Basmann, R. L. (1958) "An Experimental Investigation of Some Small-Sample Properties of (GCL) Estimators of Structural Equations: Some Preliminary Results", Richland Washington, General Electric Company, Hanford Laboratories Operation, November 21, (mimeo).

Cragg, J. G. (1967) "On the Relative Small-Sample Properties of Several Structural-Equation Estimators", Econometrica, Vol. 35, No. 1 (Jan.), 89-110.

Goldberger, A. S. (1964) "Econometric Theory", New York: John Wiley and Sons.

Kadane, J. (1971) "Comparison of k-class Estimators When the Disturbances are Small", Econometrica, Vol. 39, No. 5 (Sept.), 723-737.

Klein, L. R. (1950) "Economic Fluctuations in the United States; 1921-1941", New York: John Wiley.

Nagar, A. L. (1959) "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations", Econometrica, (Oct.), 575-595.

Sawa, T. (1973) "Almost Unibased Estimator in Simultaneous Equations Systems", International Economic Review, Vol. 14, No. 1 (Feb.), 97-106.

Sawa, T. (1973) "Foundation of Quantitative Analysis", (in Japanese) Chikuma-Shobo.

Stein, Charles (1960) "Multiple Regression", Contributions to Probability and Statistics, edited by Olkin and others, Stanford University Press.

Summers, R. (1965) "A Capital Intensive Approach to the Small Sample Properties of VArious Simultaneous Equation Estimators", Econometrica, Vol. 33, No. 1 (Jan.), 1-41.